24.4 A Leakage Current Replica Keeper for Dynamic Circuits

Yolin Lih, Nestoras Tzartzanis, William W. Walker

Fujitsu Laboratories of America, Sunnyvale, CA

Dynamic circuits are indispensable for constructing wide high-speed OR and AND-OR gates in CMOS (Fig. 24.4.1). These gates need keepers to maintain a high state during evaluation. Conventionally, a keeper is a small PFET pull-up transistor that, in the slow-PFET/fast-NFET (sPfN) process corner, sources enough current to overpower the NFET logic stack leakage current and, in the fast-PFET/slow-NFET (fPsN) process corner, is weak enough so that a single NFET leg can pull the dynamic node quickly through the switching threshold of the succeeding static gate. These contradictory requirements lead to an upper limit on the number of pull-down legs in a dynamic gate, a number that decreases with each process generation due to the exponential increase in leakage current as threshold voltages (V_t) are scaled, until eventually useful domino gates will no longer be designable [1].

Several solutions to the dynamic gate scaling problem have been proposed. In [2], a keeper is temporarily disabled using a minimum of 9 extra transistors. Additional keepers can be conditionally enabled based on a circuit that estimates the process corner [3] or the temperature [4]. None of the proposed solutions track the two critical process corners: fPsN and sPfN. In this paper, one of the most basic analog circuits, the simple current mirror, is used to replicate in series with the keeper PFET the leakage current of a reference pull-down stack (Fig. 24.4.2, note the optional *ENB* input for burn-in). The leakage current is thus faithfully reproduced in all process corners, as well as over temperature and V_{dd}. Moreover, a leakage current replica (LCR) keeper has lower overhead than other proposed solutions. Only random on-die process variations are not tracked; but after margining for these variations, LCR still has an advantage.

Figure 24.4.2 is studied in detail in what follows. For analysis, inputs B_0... B_n are first tied to high, because that is the worst case for leakage. Then the leakage current is replicated through transistors *n0...nn* by building a current mirror with a single transistor, *nrpl*, with its width equal to the sum of the widths of *n0...nn* times a safety factor, *sf*. In practice, multiple fingers are used for *nrpl* to reproduce the narrow width effect of *n0...nn*. The gate of *nrpl* is tied to V_{SS}. Transistor *p3* then mirrors the replica leakage current I_{leak} to *p1*, which is in series with *p2*. The size of *p2* is not critical, as long as it is bigger than *p1*. Since the replica current mirror can be shared with all dynamic gates having the same topology, the overhead per gate is *p1* plus a portion of the current mirror. With *sf* = 1, the current mirror will hold *DN* only to the same voltage as *KPR*. Therefore, a fairly large transistor *p3* relative to *nrpl* is first chosen to keep V(*KPR*) high, and the *sf* is selected to be approximately 10. *DN* is then pulled up first to V(*KPR*) + V_t*(p3)* where *p1* exits saturation, and continues to rise until the triode-region current of *p1* matches the actual leakage. Using high-V_t transistors, if available, for *p1* and *p3* improves the high level. Note also that if the PFET subthreshold slope is 100mV/decade, *sf* = 10 is needed to protect against 100mV V_t variation between *p1* and *p3*. Further deviation from a perfect current mirror is caused by the DIBL effect and imperfect replication of gate-tunneling leakage current, because V_{ds}*(n0...nn)* ≈ V_{dd}, whereas V_{ds}*(nrpl)* = *V(KPR)*. Both can be margined against by increasing *sf*.

The LCR keeper is evaluated and compared with a conventional keeper in a 90nm 1.2V CMOS process. The target circuit is a wide AND-OR with W=1μm pull-down NFETs (Figs. 24.4.1, 24.4.2). Inputs B_0... B_n are tied to V_{dd}, and conventional and LCR circuits are sized for a maximum voltage drop of 0.1·V_{dd} on node *DN* in the sPfN process corner with a dc noise of 0.15·V_{dd} applied to 1/4 of the A_i inputs during evaluation (*CLK*=V_{dd}). Then the gate delay is simulated in the fPsN corner with one leg active. The delay with no keeper is also simulated to show an asymptote. For LCR, node *KPR* varied from 0.829V (nominal), to 0.998 (fPsN), to 0.544 (sPfN). Figure 24.4.3 plots the delay versus the number of legs. For the conventional keeper, fPsN gives the longest delay, and the gate fails to switch with more than 24 legs. For the LCR keeper, when the number of legs is 16 or less, fPsN gives the longest delay, otherwise, sPfN is slower due to the replication of the leakage current by the keeper. With 16 to 24 legs, LCR is 25 to 40% faster than an AND-OR using a conventional keeper. Moreover, although the conventional keeper fails above 24 legs, the LCR is still usable at 32 legs.

A 72×1024 3W/4R SRAM is designed and fabricated in the same 90nm process using LCR keepers in its single-ended dynamic read path (Fig. 24.4.4). Due to dummy metal above the SRAM macro, only the IO cells and bond pads can be seen in the micrograph, so the micrograph is overlaid with a block diagram of the SRAM, PLL, and BIST circuits from the layout database. The SRAM size is 1.34 × 1.31mm^2. Previous SRAMs in this technology are designed using high-V_t NFETs in the dynamic gates, but application of LCR allowed us to switch to low-V_t NFETs to improve the speed. Figure 24.4.5 shows the 3-stage domino read path from read wordline, *RWL*, to latched output, *RDO*, which is entirely conventional except for the LCR keepers. Since all read paths are identical, only one current mirror is used for the entire SRAM, and it is placed under the wiring channels in the center (Fig. 24.4.4), giving a net overhead of one FET per dynamic gate. Due to the differing number of legs and different NFET sizes used in the pull-down stages, ratio (between W_{p1} and W_{p3} in Fig. 24.4.2, by subtracting fingers from W_{p1}) is used to divide the replica current. 2× minimum channel length is used for *p1* and *p3* to increase V_t and reduce its variation. An aggressive *sf*=6 is also used. For simulation comparison only, a design with a conventional dynamic read path is created. In the conventional design, the *KPR*-gated transistors in Fig. 24.4.5 are eliminated, all dynamic gate FETs are converted to high-V_t, and the keeper FETs were sized appropriately. Figure 24.4.6 compares simulated delays. The delay from *CLK* to *RWL* is identical for both circuits because they use identical static logic, but the delay from *RWL* to *RDO* improves by 19%, and the overall access time improves by 7.6% for the LCR SRAM.

LCR SRAMs in three process corners are fabricated and measured. Room-temperature SRAM cycle-time Shmoo plots for nominal, fPsN and sPfN process corners are compared in Fig. 24.4.7. By using the LCR keeper, all three corners perform within 20MHz (4%) of each other at the nominal 1.2V supply, in agreement with the 16-leg AO delay shown in Fig. 24.4.3.

References:

[1] M. Anders et al., "Robustness of sub-70nm Dynamic Circuits: Analytical Techniques and Scaling Trends," *Symp. VLSI Circuits*, pp. 23-24, June, 2001.

[2] A. Alvandpour et al., "A Conditional Keeper Technique for Sub-0.13μm wide Dynamic Gates," *Symp. VLSI Circuits*, pp. 29-30, 2001.

[3] C. R. Gauthier et al., US Patents #6,914,452 and #6,894,528, 2002.

[4] D. Sharishav et al., US Patent #6,791,364, 2003.

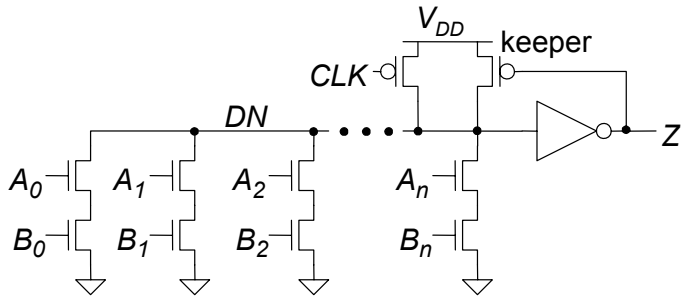

Figure 24.4.1: Generalized conventional dynamic gate topology.

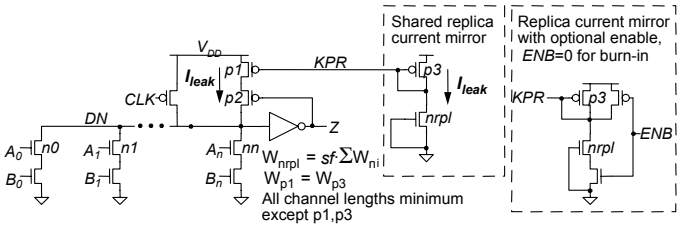

Figure 24.4.2: Proposed Leakage Current Replica keeper dynamic gate topology.

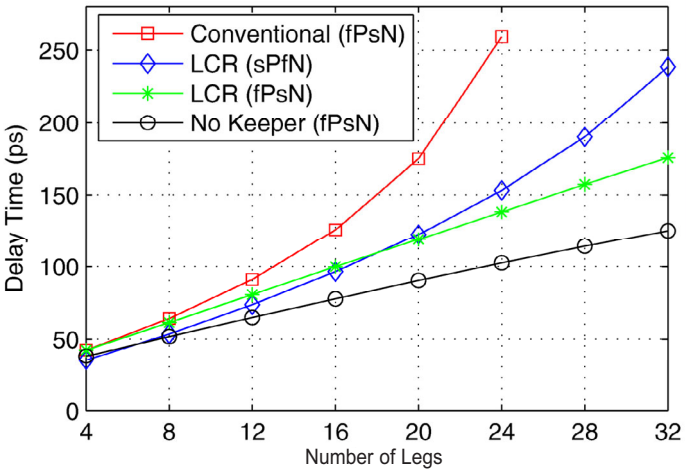

Figure 24.4.3: Delay versus number of legs.

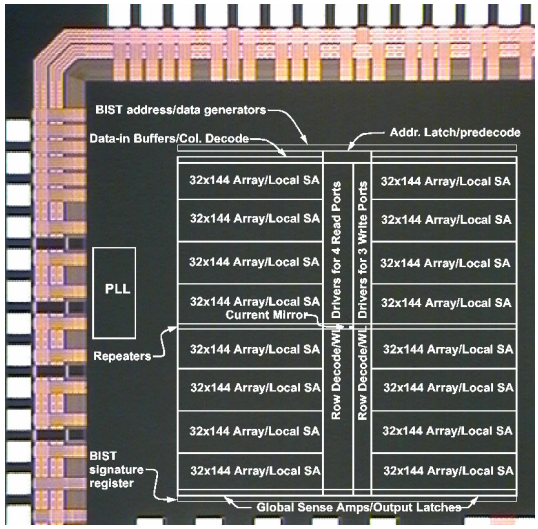

Figure 24.4.4: Micrograph of the 1024x72 3W/4R SRAM macro with block diagram overlay.

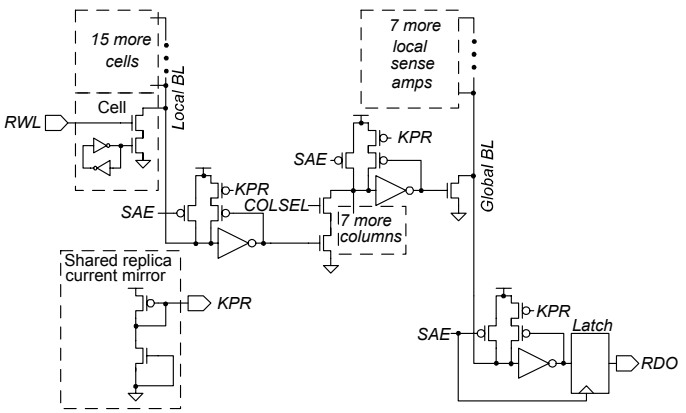

Figure 24.4.5: LCR SRAM single-ended 3-stage domino read path.

Delay Path	Conventional	LCR	Change
CLK→RWL	631ps		same
RWL→*RDO*	415ps	336ps	-19%
CLK→RDO	1046ps	967ps	-7.6%

Figure 24.4.6: Simulated SRAM access time, conventional versus LCR (typical process 1.2V, 125C).

Continued on Page 664

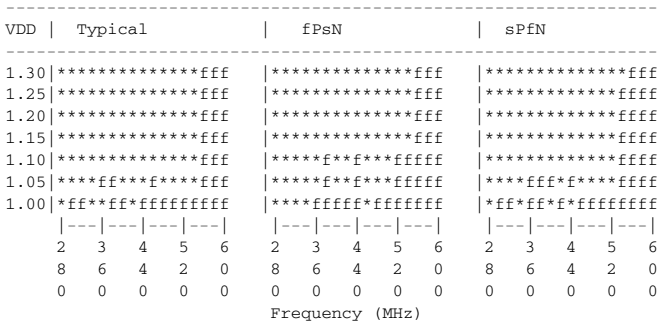

Figure 24.4.7: LCR SRAM frequency Shmoo plots (* = pass).

